PB-235 124

A MATHEMATICAL ANALYSIS OF THE KINETICS OF VIRAL INACTIVATION

Robert M. Clark, et al

National Environmental Research Center
Cincinnati, Ohio

August 1974

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-670/2-74-067	2.
3. PB 235 124	
4. TITLE AND SUBTITLE: A MATHEMATICAL ANALYSIS OF THE KINETICS OF VIRAL INACTIVATION	5. REPORT DATE: August 1974; Issuing Date
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S): Robert M. Clark, Betty Lou Grupenhoff, and George C. Kent	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS: National Environmental Research Center, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, Ohio 45268	10. PROGRAM ELEMENT NO. 1CB047; ROAP 21AQE; Task 10
	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS: Same as above	13. TYPE OF REPORT AND PERIOD COVERED
	14. SPONSORING AGENCY CODE
15. SUPPLEMENTARY NOTES	

16. ABSTRACT

Pathogenic enteric viruses transmitted via the water route present a potential hazard to public health because of their resistance to natural or artificial disinfection mechanisms. Of constant concern to public health officials is the ability of viruses to pass through water treatment plants. Therefore, many research investigations have been directed toward the study of the inactivation of viruses and enteric organisms. This report describes a mathematical model which can be used to characterize the response of viruses to a disinfecting agent. Not only is the model presented, but a technique is described which can be used to estimate the model's parameters. Both the model and the estimation technique are being used to analyze experimental information resulting from disinfection studies.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Computation, Computers, *Mathematical models, *Viruses, Linear regression, Disinfection, *Electron microscopy	Exponential decay, *Inactivation	12A, 13B

18. DISTRIBUTION STATEMENT: RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)*: UNCLASSIFIED	21. NO. OF PAGES: 40
	20. SECURITY CLASS *(This page)*: UNCLASSIFIED	22. PRICE: 3.25 - 1.45

EPA Form 2220-1 (9-73)

☆ U.S. GOVERNMENT PRINTING OFFICE: 1974— 657-049/1025

EPA-670/2-74-067
August 1974

A MATHEMATICAL ANALYSIS OF THE KINETICS OF VIRAL INACTIVATION

By

Robert M. Clark
Betty Lou Grupenhoff
George C. Kent

Water Supply Research Laboratory

Program Element No. 1CB047

NATIONAL ENVIRONMENTAL RESEARCH CENTER
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CINCINNATI, OHIO 45268

REVIEW NOTICE

This report has been reviewed by the National Environmental Research Center, Cincinnati, and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

FOREWORD

Man and his environment must be protected from the adverse effects of pesticides, radiation, noise and other forms of pollution, and the unwise management of solid waste. Efforts to protect the environment require a focus that recognizes the interplay between the components of our physical environment--air, water, and land. The National Environmental Research Centers provide this multidisciplinary focus through programs engaged in

- studies on the effects of environmental contaminants on man and the biosphere, and
- a search for ways to prevent contamination and to recycle valuable resources.

This report describes a mathematical model which can be used to characterize the response of viruses to a disinfecting agent. Not only is the model itself presented, but a technique is described which can be used to estimate the model's parameters. Both the model and the estimation technique are being used to analyze experimental information resulting from disinfection studies.

A. W. Breidenbach, Ph.D.
Director
National Environmental
Research Center, Cincinnati

ACKNOWLEDGEMENTS

Dr. P. V. Scarpino, Professor of Environmental Engineering, University of Cincinnati, provided guidance and assistance throughout all phases of this work.

Mr. Arthur F. Hammonds of the Water Supply Research Laboratory, NERC-Cincinnati, EPA, and Mr. Richard L. Manning, Office of Water Programs, Office of Air and Water Programs, EPA, Washington, D. C., assisted in the data processing work presented in this paper.

Miss Jacqueline E. A. Kent assisted in the development of and programming of several of the equations utilized in this analysis.

Ms. Catherine Hall, University of Cincinnati, assisted in the preparation of this manuscript.

Miss Gruppenhoff is employed as an engineer by General Electric Company in Cincinnati; Mr. Kent is employed by the Office of Water Programs, Office of Air and Water Programs, EPA, Washington, D. C.

A MATHEMATICAL ANALYSIS OF THE KINETICS OF VIRAL INACTIVATION

INTRODUCTION

Pathogenic enteric viruses transmitted via the water route present a potential hazard to public health because of their resistance to natural or artificial disinfection mechanisms. More than 100 different strains of enteric viruses, causing such diseases as poliomyelitis, meningitis, jaundice, and gastroenteritis, are excreted in human feces. The six major groups of enteroviruses responsible for these diseases are polioviruses, coxsackieviruses A and B, echoviruses, adenoviruses, infectious hepatitis, and reoviruses. Since these viruses are able to survive in sewage, natural waters, and water supplies, they may pose a health threat, particularly as wastewater reuse becomes more common.[1, 2]

Of constant concern to public health officials is the ability of viruses to pass through water treatment plants. The chlorine levels must be adequate, not only for bacterial disinfection but for viral inactivation as well. As a result of the need for constant concern over proper disinfection levels, much research effort has been devoted to the study of the basic disinfection mechanisms.

Chick was probably the first investigator to attempt to understand the laws of disinfection by applying the principles of first-order kinetics to bacteria and spore inactivation.[3] Only the experiments with anthrax spores conformed to first-order kinetics, whereas bacteria apparently followed another pattern of inactivation. Subsequent studies have obtained results that confirmed first-order kinetic inactivation for bacteria.

Many research investigations have been directed toward the study of the inactivation of viruses and enteric organisms. As a result of these studies, the process of inactivation has been found to be dependent on the time of contact between the organisms and disinfecting agent, concentration of disinfecting agent, temperature, and pH. In addition, viruses may form clumps of varying sizes and may cause aberrations due to their existence in inactivation systems.[4] One approach to studying the interaction of these various factors is to develop a kinetic model that will systematically account for them. The development of such a model and its application are discussed in this paper.

MODEL DEVELOPMENT

One of the major features in this model is the consideration of clumping or aggregation and its effect in explaining the devitalization process and associated aberrations. For purposes of this model, it is assumed that the virions exist either as individual particles in a suspension or as aggregates or clumps made up of two or more particles.[5] Each individual particle or aggregate will form a plaque-forming unit (PFU) before the viral suspension is subjected to a disinfecting agent. It is impossible to determine whether a PFU represents a single infective unit. If the suspension contains single particles as well as clumps of various sizes, the disinfection process will continue until the last particle in the largest clump is devitalized. When the clump is completely devitalized, a PFU is destroyed, but it is obvious that a distribution of different size clumps will lead to a non-uniform destruction of PFU's thereby causing some unusual shapes in the disinfection curve.

In this discussion, it will be assumed that this distribution of infective units represents the state of the suspension. The percentage of aggregates or clumps of all sizes which have been disinfected at any time represents the Nth state; the percentage of undisinfected single particles represents the N-1st state, etc. For illustrative purposes, let us assume a suspension in which the maximum clump size is composed of three viral particles and with clumps composed of two particles as well as single particles. Following our convention, state 1 is the percentage of undisinfected aggregates with three virions; state 2, the percentage of undisinfected aggregates with two virions; state 3, the percentage of undisinfected single particles; and state 4, the total percentage of aggregates (clumps of 1, 2, and 3 viral particles) that have been devitalized at any point in time. Obviously, under the action of a disinfectant, assuming ideal conditions, state 4 would increase as the process continues until state 4 would be 100 percent.

We can impose a frequency distribution on the various states in effect, assigning a percentage of the total plaque-forming capability to each state. The initial condition of state 4 (S_4^0) must equal 0 percent at time equal to zero or before the disinfectant acts. The percentage of undisinfected singles plus the percentages of clumps with two particles plus the percentage of clumps with three particles would equal 100 percent when time equals zero.

Associated with each state is a decay rate, k_i, that represents the probability of interaction of the destructive agent with the undisinfected singles or aggregate. The process of devitalization is assumed to take place in the following manner: The

clumps of three virions are reduced to two surviving virions, and the clumps of two are reduced to one surviving virion all the way along the chain of states until the clumps are no longer infective and are registered as a decrease in total PFU.

The set of differential equations that describes the devitalization process, where $S_i (i = 1 \ldots 4)$, the percent of plaque-forming capability at each state is:

$$\frac{dS_1}{dt} = - k_1 S_1$$

$$\frac{dS_2}{dt} = k_1 S_1 - k_2 S_2$$

$$\frac{dS_3}{dt} = k_2 S_2 - k_3 S_3$$

$$\frac{dS_4}{dt} = k_3 S_3 \qquad (1)$$

These are a set of linear first-order differential equations. The parameters k_i $(i = 1 \ldots 4)$ represents the devitalization rate with $k_4 = 0$, and S_i^0 is the initial condition of state i with $S_4^0 = 0$ at $t = 0$.

The solution to Equation 1 is as follows:

$$S_4 = k_1 k_2 k_3 S_1^0 \left[\frac{e^{-k_1 t}}{(-k_1)(k_3-k_1)(k_2-k_1)} + \frac{e^{-k_2 t}}{(-k_2)(k_3-k_2)(k_1-k_2)} \right.$$

$$\left. + \frac{e^{-k_3 t}}{(-k_3)(k_2-k_3)(k_1-k_3)} + \frac{1}{(k_3)(k_2)(k_1)} \right]$$

$$+ k_2 k_3 S_2^0 \left[\frac{e^{-k_2 t}}{(-k_2)(k_3-k_2)} + \frac{e^{-k_3 t}}{(-k_3)(k_2-k_3)} + \frac{1}{(k_3)(k_2)} \right]$$

$$- S_3^0 \left[e^{-k_3 t} - 1 \right]. \qquad (2)$$

The general closed form solution to a set of differential equations as illustrated by Equation 1 is given by the following:[5]

$$S_i = \sum_{j=1}^{i} \sum_{n=1}^{i} \frac{(k_j \cdot k_{j+1} \cdots k_{i-1})(e^{-k_n t})(S_j^0)}{(k_i - k_k)(k_{i-1} - k_k) \cdots (k_j - k_k)} \qquad (3)$$

where $k>j$. When $j=k$, $(k_j - k_k) = 1$.

When expressed as percent survival, the equation could be written as percent survival = 100 - S_i, where S_i is the last or final state to be considered.

Figure 1 illustrates schematically the change taking place during an experiment. Devitalized virus in an aggregate are represented by a broken circle. In a devitalization chain, the value for k_i, which indicates the rate of transition from one state into the next, differs for each state.[6] There are also differences between chains. For example, k_3 in the first chain may be smaller than k_3 in the second chain. This might be attributed to different geometric configurations and resulting interferences. We will assume, however, that k_3 is an average reaction rate for state 3 in all of the decay chains.

Equation 2 can be reformulated in the following manner:

$$S_4 = C_0 + C_1 e^{-k_1 t} + C_2 e^{-k_2 t} + C_3 e^{-k_3 t} \qquad (4)$$

where,

$$C_1 = \frac{-k_2 k_3 S_1^0}{(k_3 - k_1)(k_2 - k_1)}$$

$$C_2 = \frac{-k_1 k_3 S_1^0}{(k_3 - k_2)(k_1 - k_2)} - \frac{k_3 S_2^0}{(k_3 - k_2)}$$

$$C_3 = \frac{-k_1 k_2 S_1^0}{(k_2 - k_3)(k_1 - k_3)} - \frac{k_2 S_2^0}{(k_2 - k_3)} - S_3^0$$

$$C_0 = S_1^0 + S_2^0 + S_3^0 \qquad (5)$$

We know that as $t \to \infty$, $S_4 \to 100$ percent; therefore, $C_0 \to 100$ percent.

Equation 4 forms the basis for the mathematical model of the kinetics of viral inactivation we wish to examine. However, to use this equation, we must be able to estimate its parameters.

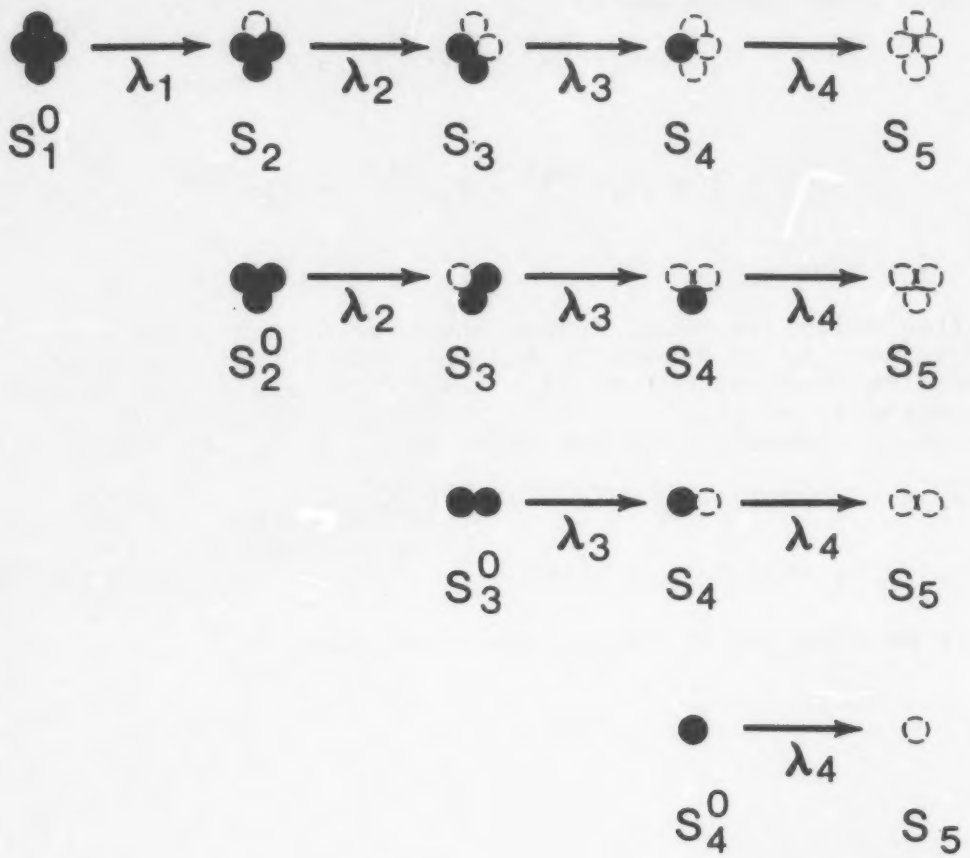

Figure 1. Schematic illustration of the physical change taking place in a suspension of virions under the influence of a devitalizing agent. Devitalized virus in a clump is represented by a broken circle. The values for λ indicate the rates of transition from one state into the next in a devitalization chain. S_i^0 is the initial percent of plaque-forming capability in state i.

ESTIMATION OF PARAMETERS

If we were to rewrite Equation 4 in terms of percent survival, we would have the following:

$$100\% - S_4^0 = -C_1e^{-k_1t} - C_2e^{-k_2t} - C_3e^{-k_3t} \tag{6}$$

or

$$\bar{y} = -C_1e^{-k_1t} - C_2e^{-k_2t} - C_3e^{-k_3t} \tag{7}$$

where $\bar{y} = 100\% - S_4^0$.

For simplicity, we shall assume that our observations are equidistant, as in Figure 2, and that the difference in the successive abscissa values is h. With the use of our three-term example, we find the value of the ith ordinate at $t_o + (i-1)h$, where t_o is the value of y_o at t_o, is then:[7]

$$\bar{y}_i = -C_1 \exp[-k_1t_0 + (i-1)h] - C_2 \exp[-k_2t_0 + (i-1)h]$$

$$-C_3 \exp[-k_3t_0 + (i-1)h] \tag{8}$$

or, if we make the following substitutions:

$$e^{-k_1h} = u_1;$$

$$e^{-k_1h} = u_2;$$

$$e^{-k_1h} = u_3;$$

$$-C_1 \exp[-k_1t_0 + (i-1)h] = f_1;$$

$$-C_2 \exp[-k_2t_0 + (i-1)h] = f_2;$$

$$-C_1 \exp[-k_3t_0 + (i-1)h] = f_3;$$

then, for five equidistant measurements, we have:

$$f_1 + f_2 + f_3 = y_i$$

$$f_1u_1 + f_2u_2 + f_3u_3 = y_{i+1}$$

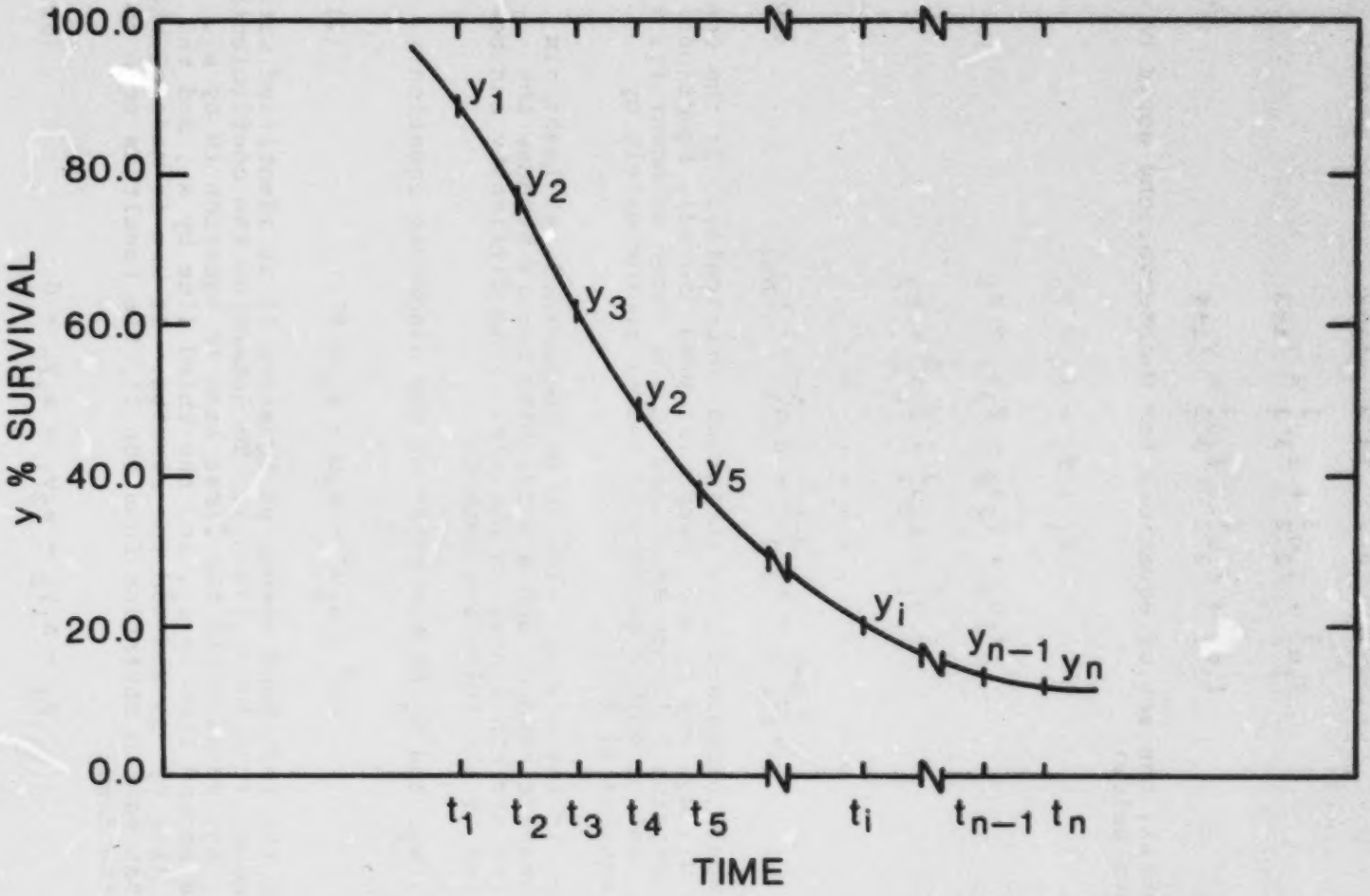

Figure 2. Equally distant values for percent survival versus time.

$$f_1u_1^2 + f_2u_2^2 + f_3u_3^2 = y_{i+2}$$

$$f_1u_1^3 + f_2u_2^3 + f_3u_3^3 = y_{i+3}$$

$$f_1u_1^4 + f_2u_2^4 + f_3u_3^4 = y_{i+4} \qquad (9)$$

In general, the set of equations for N observations would be as shown below:

$$f_1 + f_2 + f_3 = y_0$$

$$f_1u_1 + f_2u_2 + f_3u_3 = y_1$$

$$f_1u_1^2 + f_2u_2^2 + f_3u_3^2 = y_2$$

. . . . . .

$$f_1u_1^{N-1} + f_2u_2^{N-1} + f_3u_3^{N-1} = y_{N-1} \qquad (10)$$

which would necessarily be satisfied identically. If the constants u_1, u_2, and u_3 were known (or preassigned), Equations 10 would comprise N linear equations in the three unknowns f_1, f_2, and f_3, and be solved exactly if N=3 or approximately by least squares if N>3.

However, if the u's are also to be determined, at least six equations are needed, and a difficulty occurs because the equations are non-linear in the u's. This difficulty can be minimized by the following method.

Let u_1, u_2, and u_3 be the roots of the algebraic equation:

$$u^3 - a_1u^2 - a_2u - a_3 = 0 \qquad (11)$$

so that the left-hand member of Equation 11 is identified with the product $(u-u_1)(u-u_2)(u-u_3)$. To determine the coefficients a_1, a_2, a_3, we multiply the first line in Equation 10 by a_3, and the second line by a_2, and the third line by a_1, and the fourth line by -1, and add the results. If use is made of the fact that each u satisfies Equation 11, the result is seen to be of the form:

$$y_3 - a_1y_2 - a_2y_1 - a_3y_0 = 0 \qquad (12)$$

A set of N-4 additional equations of similar type is obtained in the same way by starting instead successively with the second, third . . . (N-3)th equations. In this way, we find that Equations 10 and 11 imply the N-3 linear equations:[8]

$$y_2a_1 + y_1a_2 + y_0a_3 = y_3$$

$$y_3a_1 + y_2a_2 + y_1a_3 = y_4$$

$$\cdots\cdots\cdots\cdots$$

$$y_{N-2}a_1 + y_{N-3}a_2 + \cdots y_{N-4}a_3 = y_{N-1} \qquad (13)$$

Since the ordinates y_k are known if N=6, this set generally can be solved directly for a_1, a_2, and a_3, or it can be solved approximately by the method of least squares if N>6.

In theory, after the a's are determined, the u's are found as the roots of Equation 11 and may be real or complex. Equation 10 then becomes linear and the f's can be determined from the first n of these equations or preferably by applying a least-squares technique applied to the entire set.

We have examined the situations in which there are only three terms to analyze in Equation 2. However, most often the situation will occur when there are n terms in the equation to be solved. This would take the form as follows:

$$S_n = C_0 + C_1e^{-k_1t} + C_2e^{-k_2t} + \cdots C_{n-1}e^{-k_{n-1}t} \qquad (14)$$

Assuming that there are N points equally spaced at t=0, 1, 2, 3 . . . N-1, and following the logic described in this paper, we get a set of equations similar to Equations 8:

$$f_1 + f_2 + f_3 + \cdots + f_n = y_0$$

$$f_1u_1 + f_2u_2 + f_3u_3 + \cdots + f_nu_n = y_1$$

$$f_1u_1^2 + f_2u_2^2 + f_3u_3^2 + \cdots + f_nu_n^2 = y_2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$f_1u_1^{N-1} + f_2u_2^{N-1} + f_3u_3^{N-1} + \cdots + f_nu_n^{N-1} = y_{N-1} \qquad (15)$$

Again, following the logic described earlier, we have the following N-n linear equations where the columns of data are labeled 1 through n+1.

(1)	(2)	(3)	(n)	(n+1)

$$y_{n-1}a_1 + y_{n-2}a_2 + y_{n-2}a_3 + \ldots + y_0a_n = y_n$$

$$y_na_1 + y_{n-1}a_2 + y_{n-2}a_3 + \ldots + y_1a_n = y_{n+1}$$

. . . . . .

$$y_{N-2}a_1 + y_{N-3}a_2 + y_{N-4}a_3 + \ldots + y_{N-n-1}a_n = y_{N-1} \qquad (16)$$

After the a's have been determined by least squares, the values for the c's can be found as roots of the following equation:

$$u^n - a_1u^{n-1} - a_2u^{n-2} - \ldots - a_{n-2}u - a_n = 0 \qquad (17)$$

And once the u's have been found, the f's can be found from Equation 15.

The application of this approach presumes that the number of terms that make up the model is known. Generally this number is unknown, and a major part of the analysis becomes the estimation of the optimum number of terms describing the disinfection process. Even if the number of terms is known, the solution to Equation 17 is often complex because of estimation errors in determining the coefficients. To make this analysis usable, we must be able to determine the number of terms (number of states) that make up the inactivation process. The following section describes a technique for estimating the number of components that "best" describe the inactivation process.

OPTIMAL NUMBER OF TERMS

To determine the proper number of terms that will describe the inactivation process, we would formulate the set of linear equations shown in Equation 16. In this set, the column labeled n+1 is the response or dependent variable, and the columns 1 through n are the independent variables. Using step-wise regression, we regress the independent variables (1 through n) against the n+1st or dependent variable.[9] As each variable is forced into the equation, a value for its coefficient is calculated. Each coefficient has an associated sign. When the signed coefficient is substituted into Equation 17, it is possible that an equation with alternating signs may result; for example, Equation 17 might look as follows:

$$a_0u^n - a_1u^{n-1} + a_2u^{n-2} - \ldots + a_{n-2}u - a_n = 0 \qquad (18)$$

According to Des Cartes' rule of signs:

> The number of positive real roots of a real albegraic equation either is equal to the number N_a of sign changes in the sequence $a_0, a_1, a_2, \ldots a_n$ of coefficients where vanishing terms are disregarded or it is less than N_a by a positive even integer.

Since the decay coefficients in Equation 14 are the positive real roots in Equation 18, we can use Des Cartes' rule to give us an indication as to the number of terms which optimally describes the inactivation process. We will assume that when the number of terms in the regression equation is one more than the number of sign changes, the optimal number of terms has been identified, and the variables in the regression equation are to be used in calculating k_n. The approach will be discussed beginning with the identification of the optimal number of terms.

We can illustrate this approach by assuming a model of three terms as follows:

$$y = 20.00e^{-0.10t} + 30.00e^{-0.30t} + 50.00e^{-0.50t} \qquad (19)$$

Table 1 (Page 12) contains values for Equation 19 which have been generated at intervals of t=0.50 to simulate a disinfection curve. Table 2 illustrates the way in which these data are organized to solve for the coefficients in Equation 17. As shown in Equation 16, a matrix of data points is established with n dependent variables. In this case, 27 independent variables have been constructed. The value of $y_1 = 100.00$ is the first value in the upper left-hand corner of the matrix, and the value $y_{27} = 5.7660$ is the first value for the dependent variable. The second value for the first independent value is $y_1 = 83.7858$, and the second value for $y_{28} = 5.4274$. This same pattern is repeated throughout the matrix.

Table 2. MATRIX OF DATA FOR REGRESSION ANALYSIS

Var 1	Var 2	. . .	Var n	. . .	Var 28
100.00	83.786				5.7660
83.786	70.648				5.4274
. . .	. . .				. . .
2.0359	1.9338				0.5203
1.9338	1.8370				0.4949

Table 1. VALUES FOR EQUATION 19 AT INTERVALS OF t=0.5

t	$e^{-0.100t}$	$e^{-0.300t}$	$e^{-0.500t}$	y	t	$e^{-0.100t}$	$e^{-0.300t}$	$e^{-0.500t}$	y	t	$e^{-0.100t}$	$e^{-0.300t}$	$e^{-0.500t}$	y
0.00	20.0000	30.0000	50.0000	100.0000	13.00	5.4506	0.6072	0.0751	6.1330	26.00	1.4854	0.0122	0.[illegible]	1.4978
0.50	19.0245	25.8212	38.9400	83.7858	13.50	5.1848	0.5226	0.0585	5.7660	26.50	1.4130	0.0105	0.[illegible]	1.4236
1.00	18.0967	22.2245	30.3265	70.6478	14.00	4.9319	0.4498	0.0455	5.4274	27.00	1.3441	0.0091	0.0000	1.3532
1.50	17.2141	19.1288	23.6183	59.9613	14.50	4.6914	0.3872	0.0355	5.1141	27.50	1.3785	0.0078	0.0000	1.2864
2.00	16.3746	16.4643	18.3939	51.2329	15.00	4.4626	0.3332	0.0276	4.8235	28.00	1.2163	0.0067	0.0000	1.2229
2.50	15.5760	14.1710	14.3252	44.0722	15.50	4.2449	0.2868	0.0215	4.5533	28.50	1.1568	0.0058	0.0000	1.1627
3.00	14.8163	12.1970	11.1565	38.1699	16.00	4.0379	0.2468	0.0167	4.3016	29.00	1.1004	0.0049	0.0000	1.1054
3.50	14.0937	10.4981	8.6887	33.2806	16.50	3.8410	0.2125	0.0130	4.0665	29.50	1.0467	0.0043	0.0000	1.0511
4.00	13.4064	9.0358	6.7667	29.2090	17.00	3.6536	0.1829	0.0101	3.8467	30.00	0.9957	0.0037	0.0000	0.9994
4.50	12.7525	7.7772	5.2699	25.7997	17.50	3.4754	0.1574	0.0079	3.6408	30.50	0.9471	0.0031	0.0000	0.9503
5.00	12.1306	6.6939	4.1042	22.9287	18.00	3.3059	0.1354	0.0061	3.4476	31.00	0.9009	0.0027	0.0000	0.9037
5.50	11.5390	5.7615	3.1963	20.4969	18.50	3.1447	0.11[illegible]	0.0048	3.2661	31.50	0.8570	0.0023	0.0000	0.8594
6.00	10.9762	4.9589	2.4893	18.4245	19.00	2.9913	0.100[illegible]	0.0037	3.0954	32.00	0.8152	0.0020	0.0000	0.8172
6.50	10.4409	4.2682	1.9387	16.6478	19.50	2.8454	0.0863	0.0029	2.9347	32.50	0.7754	0.0017	0.0000	0.7772
7.00	9.9317	3.6736	1.5098	15.1152	20.00	2.7067	0.0743	0.0022	2.7833	33.00	0.7376	0.0015	0.0000	0.7391
7.50	9.4473	3.1619	1.1758	13.7852	20.50	2.5747	0.0640	0.0017	2.6404	33.50	0.7016	0.0012	0.0000	0.7029
8.00	8.9865	2.7215	0.9157	12.6239	21.00	2.4491	0.0550	0.0013	2.5055	34.00	0.6674	0.0011	0.0000	0.6685
8.50	8.5483	2.3424	0.7132	11.6039	21.50	2.3296	0.0474	0.0010	2.3781	34.50	0.6349	0.0009	0.0000	0.6358
9.00	8.1313	2.0161	0.5554	10.7030	22.00	2.2160	0.0408	0.0008	2.2577	35.00	0.6039	0.0008	0.0000	0.6047
9.50	7.7348	1.7353	0.4325	9.9027	22.50	2.1079	0.0351	0.0006	2.1437	35.50	0.5744	0.0007	0.0000	0.5752
10.00	7.3575	1.4936	0.3368	9.1881	23.00	2.0051	0.0202	0.0005	2.0359	36.00	0.5464	0.0006	0.0000	0.5470
10.50	6.9987	1.2855	0.2623	8.5467	23.50	1.9073	0.0260	0.0003	1.9338	36.50	0.5198	0.0005	0.0000	0.5203
11.00	6.6574	1.1064	0.2043	7.9682	24.00	1.8143	0.0223	0.0003	1.8370	37.00	0.4944	0.0004	0.0000	0.4949
11.50	6.3327	0.9523	0.1591	7.4443	24.50	1.7258	0.0192	0.0002	1.7453					
12.00	6.0238	0.8197	0.1239	6.9675	25.00	1.6417	0.0165	0.0001	1.6584					
12.50	5.7301	0.7055	0.0965	6.5321	25.50	1.5616	0.0142	0.0001	1.5760					

Table 3 contains the results of the application of the stepwise regression program to the matrix of data in Table 2. The equations resulting from each step are as follows:

$$x^{28} - 0.9418x^{27} = 0 \quad (20)$$

$$x^{28} - 1.1223x^{27} + 0.1402x^{23} = 0 \quad (21)$$

$$x^{28} - 1.1420x^{27} + 0.1565x^{23} - 0.000089x^{1} = 0 \quad (22)$$

$$x^{28} + 0.0318x^{27} - 0.9398x^{23} + 0.07968x^{8} - 0.01124x^{1} = 0 \quad (23)$$

Equation 22 combines the maximum number of sign changes with the minimum number of variables in the equation and is, therefore, selected as the equation governing the number of terms in the disinfection equation. This matches identically with the three terms used in the simulated data. After the best estimate has been made of the number of terms which makes up the data, the next step in the analysis is to estimate the decay coefficients in the equation. This step is described in the following section.

Table 3. RESULTS OF REGRESSION ANALYSIS USING DATA FROM TABLE 2

Step	Var	Coefficient
1	27	0.94179753
2	23	-0.14022224
	27	1.12226027
3	1	0.00008941
	23	-0.15649791
	27	1.14201651
4	1	0.01124259
	8	-0.07967716
	23	0.93985016
	27	-0.03183525

ESTIMATION OF PARAMETERS

Decay Rates

Based on the data in Table 1, the parameters for Equation 19 can be estimated using the techniques outlined in Appendix A. The first decay coefficient to be calculated will be that associated with variable x^{23} and the calculation is as follows:

$$\frac{[(x^{24} - r_1 x^{23}) - (x^2 - r_1 x^1)]}{(23 - 1)} =$$

$$- \frac{[(x^{24} - r_1 x^{23}) - (x^{28} - r_1 x^{27})]}{(23 - 27)} \quad (24)$$

Substituting the average values for x^{28}, x^{27}, x^{24}, x^{23}, x^2, and x^1, into Equation 24 yields the following values for r:

$$r_1 = 0.85 \quad (25)$$

where

$$r_1 = e^{-k_1 h} \quad (26)$$

From Equation 26, we can calculate k_1 as follows:

$$k_1 = - \ln(r_1)/h \quad (27)$$

$$k_1 = - [\ln(0.85)]/0.50$$

$$k_1 = 0.32$$

The decay coefficient associated with variable x^1 is calculated as follows:

$$\ln(r_2) = \ln(x^2) - \ln(2x^1 - x^2) \quad (28)$$

$$\ln(r_1) = - 0.24$$

Substituting into Equation 27 for k_2 we get the following:

$$k_2 = (0.24)/0.50 \quad (29)$$

$$k_2 = 0.48$$

The decay coefficient associated with variable x^{27} is calculated as follows:

$$\ln(r_3) = \ln(x^{28}/x^{27}) \quad (30)$$

$$\ln(r_3) = \ln(0.95)$$

Substituting into Equation 27 for k_3 we get the following:

$$k_3 = (0.05)/0.50$$

$$k_3 = 0.10$$

Coefficients

Once the decay rates in Equation 19 have been estimated, the values for the coefficients are relatively easy to obtain. Values for each exponential term can be caluclated at the appropriate time interval and these values regressed against the values of y in Table 1. Stepwise regression can then be used to estimate the coefficients (Appendix D).

EXAMPLE INACTIVATION PROBLEM

To illustrate the utilization of this technique, it will be applied to experimental data collected from a series of electromicroscopy investigations conducted by Gordon Sharp at the University of North Carolina.[10] Sharp prepared electron micrographs of dilute preparations of T7 virus that had been subjected to a devitalizing agent.

Figure 3 shows the inactivation curve, and Table 4 contains the distribution of T7 coliphage particles resulting from these experiments. Column 1 of Table 4 lists the group size of the aggregates, that is, the number of particles in each clump of virus. Column 2 lists the number of groups in the suspension, and Column 3 lists the number of particles in each group. Column 4 lists the percent of plaque-forming capability that each group represents in the suspension. For example, there are 770 groups in the suspension, but 610/770 or 79.1 percent of them are groups of single viral particles, and 116/770 or approximately 15.1 percent of them are groups of two viral particles, etc.

Table 4. T7 VIRUS DATA

Group size	Number of groups	Number of particles	Plaque-forming capability (%)
1	610	610	79.22
2	116	232	15.06
3	24	72	3.12
4	12	48	1.56
5	6	30	0.78
6	1	6	0.13
18	1	18	0.13
Total	770	1,016	100.00

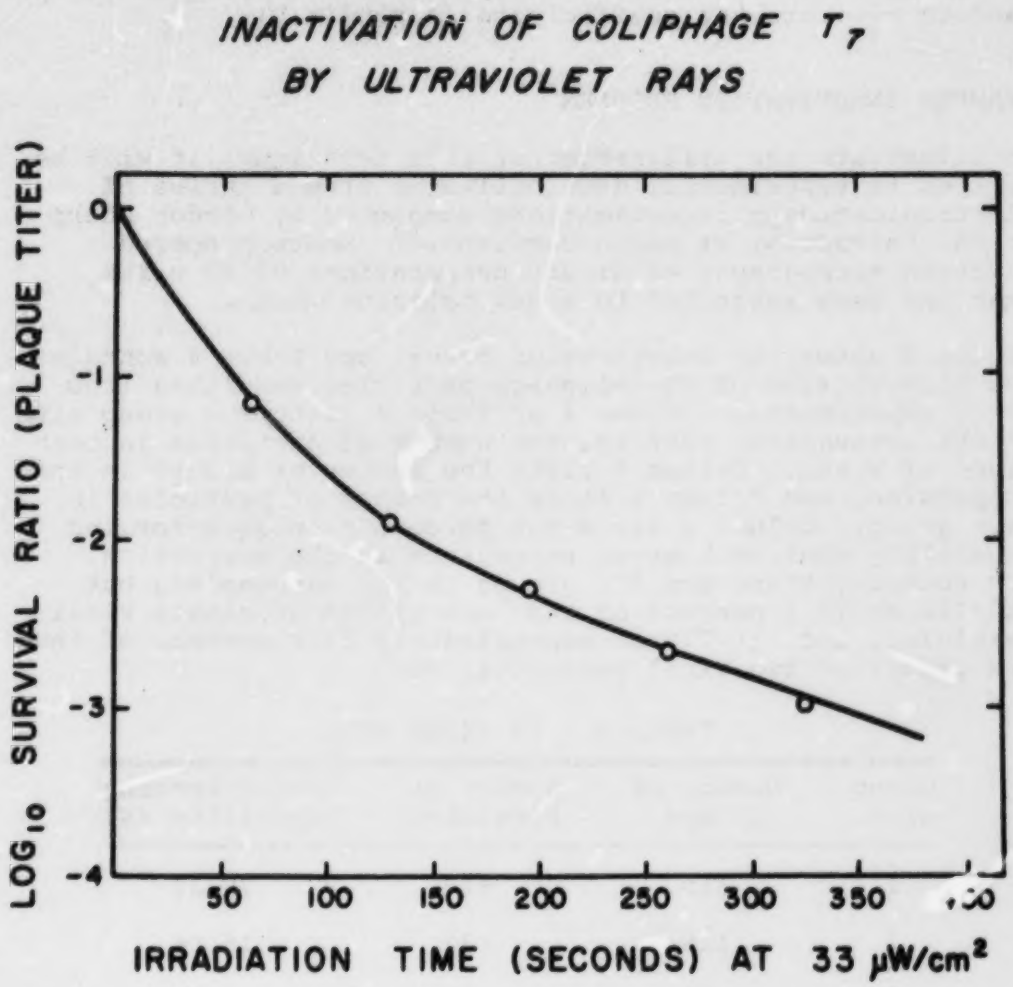

Figure 3. Inactivation of coliphage

Table 5 contains the data from Figure 3, at intervals of 5 seconds, arranged in 10 columns of data. Table 6 contains the coefficients associated with each set of variables as they enter the stepwise regression equation. It is obvious from the alternating signs that six variables will describe the inactivation process. The decay coefficients per minute calculated from the techniques outlined in Appendix A are as follows:

$$r_1 = 1.56 \qquad (32)$$

$$r_2 = 2.12$$

$$r_3 = 2.32$$

$$r_4 = 2.68$$

$$r_5 = 2.83$$

$$r_6 = 4.68$$

Each of these values represents a k_i in Equation 14, and each value of $e^{-k_i t}$ can be generated at various intervals of t by the program in Appendix B. If all six values of $e^{-k_i t}$ represented by Equation 32 are regressed against the values for y as obtained from the graph in Figure 3 then Table 7 contains the values for their coefficients. Using the program in Appendix C, the values for each predicted S_i (percentage of plaque-forming capability) can be calculated. The predicted and actual values are shown in Table 8.

When the regression is performed, the values shown in Table 7 result. At the fourth step of the regression, the corrected R^2 begins to decrease which is an indicator that the regression should be terminated at that point, and step 3 is, therefore, used as the last step in the regression analysis. Equation 5 and the program contained in Appendix C, where $k_1 = 1.56$, $k_2 = 2.83$, and $k_3 = 4.68$, yields the following values for S_i: $S_3 = 73.97\%$, $S_2 = 15.51\%$, and $S_1 = 11.30\%$. Physically, this means that there are 73.97% singles, 15.51% doubles, and the rest of the particles amount to approximately 11.30%. The comparison between the results obtained from the model and the electron micrographs is shown in Table 8. The agreement seems reasonable.

Table 5. DATA FROM DISINFECTION CURVE

Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9	Var 10
100.00000	79.00000	58.00000	45.00000	36.50000	28.50000	25.00000	19.80000	17.50000	15.00000
79.00000	58.00000	45.00000	36.00000	28.50000	25.00000	19.80000	17.50000	15.00000	13.00000
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
000.16500	00.16000	00.15500	00.15000	00.14000	00.13500	00.13000	00.12000	00.11000	00.10500
000.16000	00.15500	00.15000	00.14000	00.13500	00.13000	00.12000	00.11000	00.10500	00.10000

Table 6. RESULTS OF REGRESSION ANALYSIS USING DATA FROM TABLE 5

Step	Var	Coefficient
1	9	0.84449432
2	6	0.16137199
	9	0.57866191
3	6	0.22635578
	8	-0.22446799
	9	0.73339338
4	5	-0.14587300
	6	0.36096394
	8	-0.35799450
	9	0.95901767
5	4	0.12430537
	5	-0.26913616
	6	0.29413744
	8	-0.36433149
	9	1.01961917
6	1	-0.03634866
	4	0.29008852
	5	-0.30158431
	6	0.22952950
	8	-0.50623862
	9	1.13438561
7	1	-0.03743081
	4	0.32109643
	5	-0.43943162
	6	0.20853501
	7	0.25010067
	8	-0.67970749
	9	1.22741919

Table 7. RESULTS OF REGRESSION USING DR. SHARP'S INACTIVATION DATA

Step	Var	Coefficient	Corrected R^2 as a percent	t value
1	2	96.77	99.501	117.26
2	2	89.08	99.549	32.05
	6	6.64		2.88
3	1	58.89	99.933	19.67
	2	4.12		0.92
	6	37.77		20.82
4	1	61.68	99.932	6.69
	2	-10.82		-0.23
	4	15.24		0.32
	6	34.70		3.55

Table 8. COMPARISON BETWEEN MODEL OUTPUT AND ELECTRON MICROGRAPHS

Group size	Percent plaque-forming capability (counted)	Percent plaque-forming capability (predicted)
1	79.22	73.97
2	15.06	15.51
3 + 4 + 5 + 6 + 18	5.82	11.30

SUMMARY AND CONCLUSIONS

The kinetics of viral inactivation have been examined from a rational point of view. A mathematical model based on the radionuclide chain decay concept was formulated and a solution technique developed that allows for estimations of the optimal number of terms in the equation and for estimating the equation's parameters. With the use of data derived from electron microscopy, the model was tested and achieved reasonable results.

Based on this discussion, it is obvious that the postulated mathematical model and its solution techniques are superior to others that have been formulated. The approach outlined in this report not only determines the number of aggregate groups in the suspension, but the values for decay coefficients as well. There are some deficiencies in this approach, however, and it is important that these be considered. The approach suggested here is statistical in nature and is, therefore, subject to experimental error in the various estimations made. More importantly, the estimates of aggregate size and concentration are blind. That is, if this approach estimates three terms as optimal, there is no way to provide information on the make-up of these aggregate groupings. They might be clumps of single, double, and triple particles, or clumps of 20, 21, and 25 particles. The values for the decay coefficients may give some insight as to clump size, but these insights are hardly sufficient. This technique must be coupled with a physical assay approach incorporating electron microscopy. A research project that combines the elements of mathematical analysis with electron microscopy is currently underway.

APPENDIX A

In this appendix, the mathematical justification for the techniques used in the section entitled "Estimation of Parameters" is developed. Table 9 contains the first three values for the individual terms which make up the variables 1, 2, 23, 24, 27, and 28, as shown in Table 2. The first variable to enter the stepwise regression equation is x^{27}, as shown in Equation 20. Looking at variables 27 and 28 in Table 9, it is obvious that the term labeled f_3 dominates variable 27 and is most highly correlated with variable 28, while the terms f_1 and f_2 in variable 27 are relatively insignificant. The next variable to enter the stepwise regression equation is variable 23, and it can be seen that terms f_2 and f_3 in variable 23 are significant but that term f_1 is insignificant, and finally in variable 1, terms f_1, f_2, and f_3, are all significant. It can be concluded from this that a variable enters the regression equation when one of the terms which comprise it is significant enough to alter the rate of change of the entering variable. Therefore, we would expect that variables would enter the regression equation with alternating signs associated with their coefficients, since the entrance of each variable into the equation signifies a significant change in the functions slope. Moreover, we would expect that the variables entering the equation with alternating signs represents the maximum change in the slope of the function with respect to the other variables in the regression equation. Using Equation 19 as an example, we would, therefore, attempt to find a u_i such that b_n in the following equation is a maximum relative to its adjacent variables:

$$f_1(u_1^{n+1} - u_i u_1^n) + f_2(u_2^{n+1} - u_i u_2^n + f_3(u_3^{n+1} - u_i u_3^n) = b_n \qquad (33)$$

or in a more simple form, we would attempt to find a u_i such that b_{23} is a maximum relative to b_1 and b_{27} in the following set of equations:

$$x^2 - u_i x^1 = b_1 \qquad (34)$$

$$x^{24} - u_i x^{23} = b_{23}$$

$$x^{28} - u_i x^{27} = b_{27}$$

Table 9. FIRST THREE VALUES FOR SELECTED VARIABLES

Var	Time	$C_1e^{-0.500t}$*	$C_2e^{-0.300t}$†	$C_3e^{-0.100t}$‡	y
1	0	50.0000	30.0000	20.0000	100.0000
	0.5	38.9400	25.8212	19.0245	83.7858
	1.0	30.3265	22.2245	18.0967	70.6478
2	0.5	38.9400	25.8212	19.0245	83.7858
	1.0	30.3265	22.2245	18.0967	70.6478
	1.5	23.6183	19.1288	17.2141	59.9613
23	11.0	0.2043	1.1064	6.6574	7.9682
	11.5	0.1591	0.9523	6.3327	7.4443
	12.0	0.1239	0.8197	6.0238	6.9675
24	11.5	0.1591	0.9523	6.3327	7.4443
	12.0	0.1239	0.8197	6.0238	6.9675
	12.5	0.0965	0.7055	5.7301	6.5321
27	13.0	0.0751	0.6072	5.4506	6.1330
	13.5	0.0585	0.5226	5.1848	5.7660
	14.0	0.0455	0.4498	4.9139	5.4274
28	13.5	0.0585	0.5226	5.1848	5.7660
	14.0	0.0455	0.4498	4.9139	5.4274
	14.5	0.0355	0.3872	4.6914	5.1141

*f_1.
†f_2.
‡f_3.

Therefore, u_i can be calculated from the following equation:

$$\frac{(x^2 - u_i x^1) - (x^{24} - u_i x^{23})}{(1 - 23)} = \frac{(x^{24} - u_i x^{23}) - (x^{28} - u_i x^{22})}{(23 - 27)} \quad (35)$$

since we know that b_n is a maximum round the point x^{23}. Table 10 confirms that this is in fact the case for variable 23.

Table 10. b_n FOR SELECTED VALUES OF u_i

$x^2 - u_2 x^1$	$x^{24} - u_2 x^{23}$	$x^{28} - u_2 x^{27}$
-2.2800	0.5839	0.4857
-1.4613	0.5583	0.4449
-0.7931	0.5000	0.4569

For variables x^1 and x^{27}, this computation is impossible since there are no variables which can be used to make a computation similar to Equation 2.

However, for the decay coefficient associated with x^1, the following equation can be developed using the properties of infinite series:

$$u_i^n(x^2 - u_i x^1) \to u_i^{n+1}(x^1 - x^2) \text{ as } n \to \infty \quad (36)$$

Therefore,

$$\ln(u_i) = \ln(x^2) - \ln(2x^1 - x^2) \quad (37)$$

For the decay coefficient associated with x^{27}, the following relationship can be developed:

$$u_i^n x^{28} - u_i^{n+1} x^{27} \to 0 \text{ as } n \to \infty \quad (38)$$

Therefore,

$$\ln(u_i) = \ln(x^{28}) - \ln(x^{27}) \quad (39)$$

APPENDIX B

PROGRAM FOR GENERATING VALUES OF $e^{-k_i t}$

AT GIVEN INTERVALS OF t

```
      DIMENSION A(15),B(15),C(15)
      MR = 1
      MW = 15
    5 READ(MR,10)T,N
   10 FORMAT(F5.2,I2)
      IF (N)70,70,15
   15 READ(MR,20)(A(J),J=1,N)
      READ(MR,20)(C(J),J=1,N)
   20 FORMAT(10F8.0)
      E=2.71828
      S=0.0
      Y=0.0
      WRITE(MW,25)
   25 FORMAT('1'//26X,'-XT')
      WRITE(MW,30)
   30 FORMAT(25X,'E     TABLE')
      WRITE(MW,35)
   35 FORMAT(25X,'-----------')
      WRITE(MW,40)(A(J),J=1,N)
   40 FORMAT(//10X, 3(3X,'-',F5.3,'T'))
      WRITE(MW,45)
   45 FORMAT(5X,'T',6X,'E', 2(9X,'E'),12X,'Y')
   50 DO 60 I=1,N
   55 B(I)=C(I)*E**(-A(I)*S)
   60 Y=Y+B(I)
      WRITE(MW,65) S,(B(K),K=1,N),Y
   65 FORMAT(/3X,F5.2,2X,11F10.4)
      S=S+T
      Y=0.0
      IF (B(1)-.0001)     5,50,50
   70 STOP
      END
```

APPENDIX C

PROGRAM FOR CALCULATION S_i

```
      DIMENSION A(16),B(16),S(10),TEMP(10),STOR(10)                    001
    2 READ (1,5)N                                                      002
    5 FORMAT (I2)                                                      003
      IF(N)8,42,8
    8 READ (1,10)A,B                                                   004
   10 FORMAT (8F10.3/8F10.3)                                           005
      DO 41 I=1,N                                                      006
      TEMP(1)=1.0                                                      006A
      STOR(1)=1.0                                                      006B
      DO 30 J=1,N                                                      007
      IF(I-J)15,30,15                                                  008
   15 TEMP(1)=TEMP(1)*(A(J)-A(I))                                      009
      STOR(1)=STOR(1)*A(J)                                             010
   30 CONTINUE                                                         011
      U1=TEMP(1)/STOR(1)                                               012
      IF(I-1)45,35,45                                                  013
   35 S(1)=B(1)*U1                                                     014
      WRITE (3,40) S(1)                                                015
   40 FORMAT ('1'////' S(1) = ',F12.2///)
      GO TO 41                                                         017
   45 TEMP(2)=1.0                                                      017A
      STOR(2)=1.0                                                      017B
      DO 65 J=1,N                                                      018
      IF(I-J)50,65,50                                                  019
   50 STOR(2)=STOR(2)*A(J)                                             020
      TEMP(2)=TEMP(2)*(A(J)-A(I))                                      021
   65 CONTINUE                                                         022
      U2=STOR(2)*S(1)/TEMP(2)                                          023
      IF(I-2)80,70,80                                                  024
   70 S(2)=(B(2)-U2)*U1                                                025
      WRITE (3,75) S(2)                                                026
   75 FORMAT (' S(2) = ',F12.2///)                                     027
      GO TO 41                                                         028
   80 TEMP(3)=1.0                                                      028A
      STOR(3)=1.00                                                     028B
      DO 100 J=2,N                                                     029
      IF(I-J)85,100,85                                                 030
   85 STOR(3)=STOR(3)*A(J)                                             031
      TEMP(3)=TEMP(3)*(A(J)-A(I))                                      032
  100 CONTINUE                                                         033
      U3=STOR(3)*S(2)/TEMP(3)                                          034
      IF(I-3)115,105,115                                               035
  105 S(3)=(B(3)-U2-U3)*U1                                             036
      WRITE (3,110) S(3)                                               037
  110 FORMAT (' S(3) = ',F12.2///)                                     038
      GO TO 41                                                         039
  115 TEMP(4)=1.0                                                      039A
      STOR(4)=1.0                                                      039B
      DO 135 J=3,N                                                     040
      IF(I-J)120,135,120                                               041
  120 STOR(4)=STOR(4)*A(J)                                             042
      TEMP(4)=TEMP(4)*(A(J)-A(I))                                      043
  135 CONTINUE                                                         044
      U4=STOR(4)*S(3)/TEMP(4)                                          045
      IF(I-4)150,140,150                                               046
  140 S(4)=(B(4)-U2-U3-U4)*U1                                          047
      WRITE (3,145) S(4)                                               048
  145 FORMAT (' S(4) = ',F12.2///)                                     049
      GO TO 41                                                         050
  150 TEMP(5)=1.0                                                      050A
      STOR(5)=1.0                                                      050B
```

```
      DO 170 J=4,N                                                    051
      IF(I-J)155,170,155                                              052
  155 STOR(5)=STOR(5)*A(J)                                            053
      TEMP(5)=TEMP(5)*(A(J)-A(I))                                     054
  170 CONTINUE                                                        055
      U5=STOR(5)*S(4)/TEMP(5)                                         056
      IF(I-5)185,175,185                                              057
  175 S(5)=(B(5)-U2-U3-U4-U5)*U1                                      058
      WRITE (3,180) S(5)                                              059
  180 FORMAT (' S(5) = ',F12.2///)                                    060
      GO TO 41                                                        061
  185 TEMP(6)=1.0                                                     061A
      STOR(6)=1.0                                                     061B
      DO 205 J=5,N                                                    062
      IF(I-J)190,205,190                                              063
  190 STOR(6)=STOR(6)*A(J)                                            064
      TEMP(6)=TEMP(6)*(A(J)-A(I))                                     065
  205 CONTINUE                                                        066
      U6=STOR(6)*S(5)/TEMP(6)                                         067
      IF(I-6)220,210,220                                              068
  210 S(6)=(B(6)-U2-U3-U4-U5-U6)*U1                                   069
      WRITE (3,215) S(6)                                              070
  215 FORMAT (' S(6) = ',F12.2///)                                    071
      GO TO 41                                                        072
  220 TEMP(7)=1.0                                                     072A
      STOR(7)=1.0                                                     072B
      DO 240 J=6,N                                                    073
      IF(I-J)225,240,225                                              074
  225 STOR(7)=STOR(7)*A(J)                                            075
      TEMP(7)=TEMP(7)*(A(J)-A(I))                                     076
  240  CONTINUE                                                       077
      U7=STOR(7)*S(6)/TEMP(7)                                         078
      IF(I-7)255,245,255                                              079
  245 S(7)=(B(7)-U2-U3-U4-U5-U6-U7)*U1                                080
      WRITE (3,250) S(7)                                              081
  250 FORMAT (' S(7) = ',F12.2///)                                    082
      GO TO 41                                                        083
  255 TEMP(8)=1.0                                                     083A
      STOR(8)=1.0                                                     083B
      DO 275 J=7,N                                                    084
      IF(I-J)260,275,260                                              085
  260 STOR(8)=STOR(8)*A(J)                                            086
      TEMP(8)=TEMP(8)*(A(J)-A(I))                                     087
  275 CONTINUE                                                        088
      U8=STOR(8)*S(7)/TEMP(8)                                         089
      IF(I-8)290,280,290                                              090
  280 S(8)=(B(8)-U2-U3-U4-U5-U6-U7-U8)*U1                             091
      WRITE (3,285) S(8)                                              092
  285 FORMAT (' S(8) = ',F12.2///)                                    093
      GO TO 41                                                        094
  290 TEMP(9)=1.0                                                     094A
      STOR(9)=1.0                                                     094B
      DO 300 J=8,N                                                    095
      IF(I-J)295,300,295                                              096
  295 STOR(9)=STOR(9)*A(J)                                            097
      TEMP(9)=TEMP(9)*(A(J)-A(I))                                     098
  300 CONTINUE                                                        099
      U9=STOR(9)*S(8)/TEMP(9)                                         100
      IF(I-9)315,305,315                                              101
  305 S(9)=(B(9)-U2-U3-U4-U5-U6-U7-U8-U9)*U1                          102
      WRITE (3,310) S(9)                                              103
```

```
310 FORMAT (' S(9) = ',F12.2///)                                    104
    GO TO 41                                                        105
315 TEMP(10)=1.0                                                    105A
    STOR(10)=1.0                                                    105B
     DO 325 J=9,N                                                   106
    IF(I-J)320,325,320                                              107
320 STOR(10)=STOR(10)*A(J)                                          108
    TEMP(10)=TEMP(10)*(A(J)-A(I))                                   109
325 CONTINUE                                                        110
    U10=STOR(10)*S(9)/TEMP(10)                                      111
    IF(I-10)41,330,41                                               112
330 S(10)=(B(10)-U2-U3-U4-U5-U6-U7-U8-U9-U10)*U1                    113
    WRITE (3,335) S(10)                                             114
335 FORMAT (' S(10) = ',F12.2///)                                   115
 41 CONTINUE                                                        116
    GO TO 2                                                         116A
 42 STOP                                                            117
    END                                                             118
```

APPENDIX D

STEPWISE REGRESSION PROGRAM

```
C      1130 STEPWISE MULTIPLE REGRESSION PROGRAM, 3/14/66                    0010
C      PHASES 1 AND 2 CAN BE OVERLAID TO CONSERVE CORE. THE STEPS TO         0020
C      READY PHASES 1 AND 2 FOR OVERLAY ARE                                  0030
C         1. SET UP A COMMON AREA CONSISTING OF RIJ,XBAR,SIGMA,FIN,          0040
C         FOUT,OBS,NVAR,NOBS,NINDV,IRES,IFA.                                 0050
C         2. SET SIGMA AND DATA EQUIVALENT IN PHASE 2.                       0060
C         3. REPEAT PHASE 1 DEFINE FILE STATEMENT IN PHASE 2.                0070
C         4. REMOVE STATEMENT 101-3 FROM PHASE 1 AND INSERT IT               0080
C         BEHIND DIMENSION COMMENTS CARD IN PHASE 2.                         0090
C      PHASE 1. TRANSFORM ORIGINAL DATA, COMPUTE AND PRINT MEANS,            0100
C      STANDARD DEVIATIONS, AND SIMPLE CORRELATION COEFFICIENTS.             0110
C      DIMENSIONS                                                            0120
       IMPLICIT REAL*8(A-H,O-Z)
       DIMENSION DATA(30),CONST(12),ITRAN(30),JTRAN(30),KTRAN(30),LTRAN(3    0130
      10)                                                                    0140
       DIMENSION RIJ(30,30),XBAR(30), SIGMA(30),AID(18)                      0150
       DIMENSION SIGB(30),B(30),ID(30)                                       0160
C      EQUIVALENCES                                                          0170
       EQUIVALENCE (SIGMA(1),DATA(1))                                        0180
C      DEFINE DATA FILE                                                      0190
       DEFINE FILE 10(1000,60,U,IFA)
C      STATEMENT LABEL 101 IS NOT REFERENCED. IT MARKS THE FIRST             0210
C      EXECUTABLE STATEMENT OF THE SOURCE PROGRAM.                           0220
C      ICOM IS FIXED DECIMAL REPRESENTATION OF ALPHABETIC COMMA.             0230
  101  ICOM=27456                                                            0240
C      INITIALIZE DATA FILE                                                  0250
       IFA=1                                                                 0260
C      READ I.D.                                                             0270
       READ(1,1, END=999) (AID(I),I=1,18)
     1 FORMAT(18A4)
C      READ CONTROL CARD                                                     0300
       READ(1,2)NVIN,NVAR,NOBS,NTRAN,NCONS,FIN,FOUT,IRES                     0310
     2 FORMAT(2I2,I4,2I2,2F6.3,I1)
       IF(FIN-FOUT)1020,690,690                                              331
   690 IF(NTRAN)1000,730,700                                                 0340
C      READ TRANSFORMATION CARDS                                             0350
   700 READ(1,71)(ITRAN(I),JTRAN(I),KTRAN(I),LTRAN(I),I=1,NTRAN)            0360
    71 FORMAT(36I2)                                                          0370
       IF(NCONS)1000,730,720                                                 0380
C      READ CONSTANT CARD                                                    0390
   720 READ(1,72)(CONST(I),I=1,NCONS)                                        0400
    72 FORMAT(12F6.3)                                                        0410
C      INITIALIZE.                                                           0420
   730 OBS=NOBS                                                              0430
       NINDV=NVAR-1                                                          0440
       DO 90 I=1,NVIN                                                        0450
       XBAR(I)=0.0
       DO 90 J=1,NVIN                                                        0470
    90 RIJ(I,J)=0.0                                                          0480
C      READ DATA, FORM SUMS VECTOR, SUMS OF SQUARES MATRIX                   0490
       DO 110 I=1,NOBS                                                       0500
       READ(1,3)(DATA(J),J=1,NVIN)                                           0510
     3 FORMAT(12F6.0)
       IF(NTRAN)1000,860,750                                                 0530
C      TRANSFORMATION OF RAW DATA                                            0540
   750 DO 850 M=1,NTRAN                                                      0550
       II=ITRAN(M)                                                           0560
       JJ=JTRAN(M)                                                           0570
       KK=KTRAN(M)                                                           0580
       LL=LTRAN(M)                                                           0590
       GO TO (760,770,780,790,800,810,820,830,840),II                        0600
```

```
C       X(J)=X(K)                                                       0610
  760 DATA(JJ)=DATA(KK)                                                 0620
        GO TO 850                                                       0630
C       X(J)=-X(K)                                                      0640
  770 DATA(JJ)=-DATA(KK)                                                0650
        GO TO 850                                                       0660
C       X(J)=LOG X(K)                                                   0670
  780 DATA(JJ)=DLOG(DATA(KK))                                           0680
        GO TO 850                                                       0690
C       X(J)=1/X(K)                                                     0700
  790 DATA(JJ)=1.0/DATA(KK)                                             0710
        GO TO 850                                                       0720
C       X(J)=X(K)+X(L)                                                  0730
  800 DATA(JJ)=DATA(KK)+DATA(LL)                                        0740
        GO TO 850                                                       0750
C       X(J)=X(K)*X(L)                                                  0760
  810 DATA(JJ)=DATA(KK)*DATA(LL)                                        0770
        GO TO 850                                                       0780
C       X(J)=X(K)/X(L)                                                  0790
  820 DATA(JJ)=DATA(KK)/DATA(LL)                                        0800
        GO TO 850                                                       0810
C       X(J)=X(K)+C(L)                                                  0820
  830 DATA(JJ)=DATA(KK)+CONST(LL)                                       0830
        GO TO 850                                                       0840
C       X(J)=X(K)*C(L)                                                  0850
  840 DATA(JJ)=DATA(KK)*CONST(LL)                                       0860
  850 CONTINUE                                                          0870
  860 IF(IRES)870,880,870                                               0880
C       WRITE DATA FILE                                                 0890
  870 WRITE(10'IFA)(DATA(J),J=1,NVAR)
  880 DO 100 J=1,NVAR                                                   0910
        XBAR(J)=XBAR(J)+DATA(J)                                         0920
        DO 100 K=1,NVAR                                                 0930
  100 RIJ(J,K)=RIJ(J,K)+DATA(J)*DATA(K)                                 0940
  110 CONTINUE                                                          0950
C       COMPUTE STANDARD DEVIATIONS*SQR ROUTE (OBS-1)                   0960
        DO 120 I=1,NVAR                                                 0970
  120 SIGMA(I)=(RIJ(I,I)-XBAR(I)*XBAR(I)/OBS)**.5                       0980
C       COMPUTE CORRELATION MATRIX                                      0990
        DO 130 I=1,NVAR                                                 1000
        DO 130 J=1,NVAR                                                 1010
  130 RIJ(I,J)=(RIJ(I,J)-XBAR(I)*XBAR(J)/OBS)/(SIGMA(I)*SIGMA(J))       1020
C       COMPUTE MEANS AND STANDARD DEVIATIONS                           1030
        DO 140 I=1,NVAR                                                 1040
        XBAR(I)=XBAR(I)/OBS                                             1050
  140 SIGMA(I)=SIGMA(I)/(OBS-1.0)**.5                                   1060
C       SKIP TO NEW PAGE, WRITE I.D., AVERAGES, STANDARD DEVIATIONS,    1070
C       AND SIMPLE CORRELATION MATRIX.                                  1080
        WRITE(3,65)(AID(I),I=1,18)                                      1090
   65 FORMAT('1',18A4)                                                  1100
        WRITE(3,51)                                                     1110
   51 FORMAT('0AVERAGES')                                               1120
        WRITE(3,52)(I,XBAR(I),ICOM,I=1,NINDV),NVAR,XBAR(NVAR)           1130
   52 FORMAT (4(' VAR(',I2,')=',F13.2,A1))                              1140
        WRITE(3,53)                                                     1150
   53 FORMAT('0STANDARD DEVIATIONS')                                    1160
        WRITE(3,52)(I,SIGMA(I),ICOM,I=1,NINDV),NVAR,SIGMA(NVAR)         1170
        WRITE(3,55)                                                     1180
   55 FORMAT('0SIMPLE CORRELATION COEFFICIENTS')                        1190
        DO 150 I=1,NINDV                                                1200
        WRITE(3,56)(I,J,RIJ(I,J),ICOM,J=I,NINDV),I,NVAR,RIJ(I,NVAR)     1210
```

```
   56 FORMAT (14(' VARS(',I2,',',I2,')=',F10.3,A1))                    1220
  150 CONTINUE
C     PHASE 2. PERFORM STEPWISE CALCULATIONS AND PRINT RESULTS.       1230
C     DIMENSIONS                                                      1240
C     INITIALIZE                                                      1250
      DO 190 I=1,NVAR                                                 1260
      SIGB(I)=0.0                                                     1270
  190 B(I)=0.0                                                        1280
      NENT=0                                                          1290
      DF=OBS-1.0                                                      1300
      NSTEP=-1                                                        1310
C     TRANSFORM SIGMA VECTOR FROM STANDARD DEVIATIONS TO SQUARE       1320
C     ROOTS OF SUMS OF SQUARES.                                       1330
      DO 310 I=1,NVAR                                                 1340
  310 SIGMA(I)=SIGMA(I)*(OBS-1.0)**.5                                 1350
C     BEGIN STEP NUMBER NSTEP.                                        1360
  200 NSTEP=NSTEP+1                                                   1370
      STDEE=((RIJ(NVAR,NVAR)/DF)**.5)*SIGMA(NVAR)                     1380
      DF=DF-1.0                                                       1390
      IF(DF)1010,1010,205                                             1400
  205 VMIN=0.0                                                        1410
      VMAX=0.0                                                        1420
      NIN=0                                                           1430
C     FIND MINIMUM VARIANCE CONTRIBUTION OF VARIABLES IN REGRESSION   1440
C     EQUATION. FIND MAXIMUM VARIANCE CONTRIBUTION OF VARIABLES       1450
C      NOT IN REGRESSION EQUATION.                                    1460
      DO 300 I=1,NINDV                                                1470
  210 VI=RIJ(I,NVAR)*RIJ(NVAR,I)/RIJ(I,I)                            1490
      IF(VI)240,300,220                                               1500
  220 IF(VI-VMAX)300,300,230                                          1510
  230 VMAX=VI                                                         1520
      NMAX=I                                                          1530
      GO TO 300                                                       1540
  240 NIN=NIN+1                                                       1550
      ID(NIN)=I                                                       1560
C     COMPUTE REGRESSION COEFFICIENT AND ITS STANDARD DEVIATION.      1570
      B(NIN)=RIJ(I,NVAR)*SIGMA(NVAR)/SIGMA(I)                         1580
      SIGB(NIN)=(STDEE*RIJ(I,I)**.5)/SIGMA(I)                         1590
      IF(VMIN)250,260,1000                                            1600
  250 IF(VI-VMIN)300,300,260                                          1610
  260 VMIN=VI                                                         1620
      NMIN=I                                                          1630
  300 CONTINUE                                                        1640
      IF(NIN)1000,460,400                                             1650
C     COMPUTE CONSTANT TERM.                                          1660
  400 BSUBO=XBAR(NVAR)                                                1670
      DO 410 I=1,NIN                                                  1680
      J=ID(I)                                                         1690
  410 BSUBO=BSUBO-B(I)*XBAR(J)                                        1700
      IF(NENT)1000,480,420                                            1710
C     OUTPUT FOR VARIABLE ADDED                                       1720
  420 WRITE(3,57)NSTEP,K                                              1730
   57 FORMAT('0STEP NUMBER ',I2,10X,'ENTER VARIABLE ',I2)             1740
      DEPV = NSTEP
  425 WRITE(3,58)STDEE                                                1750
   58 FORMAT(' STANDARD DEVIATION OF RESIDUALS=',F16.3)               1760
      SDPRM=(STDEE/XBAR(NVAR))*100.                                   1761
      WRITE(3,49)SDPRM                                                1762
   49 FORMAT(' STD. DEV. AS PERCENT OF RESPONSE MEAN=',F10.3)         1763
      R=(1.-RIJ(NVAR,NVAR))**.5                                       1770
      RSQ = R**2.
```

```
      RSQP = RSQ * 100.
      WRITE(3,59) RSQP
   59 FORMAT(' PERCENT VARIATION EXPLAINED R-SQ=',F15.3)                    1773
      CRSQ = 1.-((1.-RSQ)*(OBS-1.))/(OBS-DEPV-1.)
      CRSQP = CRSQ * 100.
      WRITE(3,84) CRSQP
   84 FORMAT(' CORRECTED R-SQ AS A PERCENT=',F20.3)
      IDFN=OBS-DF-2.0                                                       1800
      IDFD=DF+1.0                                                           1810
      F=(SIGMA(NVAR)**2-(STDEE**2)*(DF+1.0))/((OBS-DF-2.0)*STDEE**2)        1820
      WRITE(3,66)IDFN,IDFD,F                                                1830
   66 FORMAT(' GOODNESS OF FIT OR OVERALL F,F(',I3,',',I3,')=',F8.3)        1840
      WRITE(3,60)BSUBO                                                      1850
   60 FORMAT(' CONSTANT TERM=',18X,F16.8)                                   1860
      WRITE(3,61)                                                           1870
   61 FORMAT('OVAR         COEFF                STD DEV             T VALUE'  1880
     1)                                                                     1881
      WRITE(3,62)                                                           1890
   62 FORMAT('                                   COEFF')                    1900
      DO 430 I=1,NIN                                                        1910
      J=ID(I)                                                               1920
      T=B(I)/SIGB(I)                                                        1930
      WRITE(3,63)ID(I),B(I),SIGB(I),T                                       1940
   63 FORMAT(' ',I3,F18.8,F20.8,F18.8)                                      1950
  430 CONTINUE
C     COMPUTE F LEVEL FOR MINIMUM VARIANCE CONTRIBUTION VARIABLE            1960
C     IN REGRESSION EQUATION.                                               1970
      FLEVL=VMIN*DF/RIJ(NVAR,NVAR)                                          1980
      IF(FOUT+FLEVL)460,460,450                                             1990
C     INITIALIZE FOR REMOVAL OF VARIABLE K FROM EQUATION.                   2000
  450 K=NMIN                                                                2010
      NENT=0                                                                2020
      DF=DF+2.0                                                             2030
      GO TO 500                                                             2040
C     COMPUTE F LEVEL FOR MAXIMUM VARIANCE CONTRIBUTION VARIABLE            2050
C     NOT IN EQUATION.                                                      2060
  460 FLEVL=VMAX*DF/(RIJ(NVAR,NVAR)-VMAX)                                   2070
      IF(FLEVL-FIN)600,600,470                                              2080
C     INITIALIZE FOR ENTRY OF VARIABLE K INTO EQUATION.                     2090
  470 K=NMAX                                                                2100
      NENT=K                                                                2110
      GO TO 500                                                             2120
C     OUTPUT FOR VARIABLE DELETED                                           2130
  480 WRITE(3,64)NSTEP,K                                                    2140
   64 FORMAT('0STEP NUMBER ',I2,10X,'DELETE VARIABLE ',I2)                  2150
      GO TO 425                                                             2160
C     UPDATE MATRIX                                                         2170
  500 DO 540 I=1,NVAR                                                       2180
      IF(I-K)510,540,510                                                    2190
  510 DO 530 J=1,NVAR                                                       2200
      IF(J-K)520,530,520                                                    2210
  520 RIJ(I,J)=RIJ(I,J)-RIJ(I,K)*RIJ(K,J)/RIJ(K,K)                          2220
  530 CONTINUE                                                              2230
  540 CONTINUE                                                              2240
      DO 560 J=1,NVAR                                                       2250
      IF(J-K)550,560,550                                                    2260
  550 RIJ(K,J)=RIJ(K,J)/RIJ(K,K)                                            2270
  560 CONTINUE                                                              2280
      DO 580 I=1,NVAR                                                       2290
      IF(I-K)570,580,570                                                    2300
  570 RIJ(I,K)=-RIJ(I,K)/RIJ(K,K)                                           2310
```

```
  580 CONTINUE                                                          2320
      RIJ(K,K)=1.0/RIJ(K,K)                                             2330
      GO TO 200                                                         2340
  600 IF(IRES)610,640,610                                               2350
C     PRINT RESIDUALS                                                   2360
  610 IFA=1                                                             2370
      WRITE(3,67)                                                       2380
   67 FORMAT('0 OBS        ACTUAL     ESTIMATE     RESIDUAL      NORMAL')2390
      WRITE(3,69)                                                       2391
   69 FORMAT('                                                DEVIATE')2392
      DO 630 K=1,NOBS                                                   2400
      READ(10'IFA)(DATA(I),I=1,NVAR)                                    2410
      EST=BSUBO                                                         2420
      DO 620 I=1,NIN                                                    2430
      J=IO(I)                                                           2440
  620 EST=EST+B(I)*DATA(J)                                              2450
      RESID = DATA(NVAR)-EST                                            2460
      XNORD = RESID/STDEE                                               2461
      IF(DABS(XNORD)-3.)91,92,92                                        2470
   91 IF(DABS(XNORD)-2.)93,94,94                                        2471
   92 WRITE(3,30)K,DATA(NVAR),EST,RESID,XNORD                           2480
   30 FORMAT(' ',I4,4F12.2,' **')                                       2481
      GO TO 630                                                         2482
   94 WRITE(3,31)K,DATA(NVAR),EST,RESID,XNORD                           2483
   31 FORMAT(' ',I4,4F12.2,' *')                                        2484
      GO TO 630                                                         2485
   93 WRITE(3,68)K,DATA(NVAR),EST,RESID,XNORD                           2486
   68 FORMAT(' ',I4,4F12.2)                                             2487
  630 CONTINUE                                                          2490
C     NORMAL END OF JOB                                                 2500
  640 GO TO 101                                                         2501
  999 CALL EXIT
C     ERROR. NIN, NENT, VMIN, NCONS, OR NTRANS IS NEGATIVE. CHECK       2520
C     FOR CONTROL CARD ERROR.                                           2530
 1000 STOP1
C     ERROR  DEGREES OF FREEDOM =0. EITHER ADD MORE DATA OBSERVATIONS OR 2550
C     DELETE ONE OR MORE INDEPENDENT VARIABLES. SAMPLE SIZE MUST EXCEED 2560
C     NUMBER OF INDEPENDENT VARIABLES BY AT LEAST 2.                    2570
 1010 STOP2                                                             2580
C     ERROR. F LEVEL FOR INCOMING VARIABLE IS LESS THAN F LEVEL FOR     2590
C     OUTGOING VARIABLE.                                                2600
 1020 STOP4                                                             2610
      END
```

REFERENCES

1. Berg, G., "Virus Transmission by the Water Vehicle," I. Viruses, Health Lab Science, 3:86 (1966).

2. Berg, G., "Virus Transmission by the Water Vehicle," III. Removal of Viruses by Water Treatment Procedures, Health Lab Science, 3:170 (1966).

3. Chick, H., "An Investigation of the Laws of Disinfection," Journal of Hygiene, 8:92 (1908).

4. Berg, G.; Clark, R. M.; Berman, D.; and Chang, S. L., "Aberrations in Survival Curves," Transmission of Viruses by the Water Route, Interscience Publishers, a division of John Wiley and Sons, New York, New York, pp. 235-240 (1967).

5. Clark, R. M., and Niehaus, J. F., "A Mathematical Model for Viral Devitalization," Transmission of Viruses by the Water Route, Interscience Publishers, a division of John Wiley and Sons, New York, New York, pp. 241-245 (1967).

6. Clark, R. M., "A Mathematical Model of the Kinetics of Viral Devitalization," Mathematical Biosciences 2, pp. 413-423 (1968).

7. Willers, A., FR, Practical Analysis, Dover, New York (1948).

8. Hildebrand, F. B., Introduction to Numerical Analysis, McGraw-Hill, New York, New York (1956).

9. Draper, N., and Smith, H., Applied Regression Analysis, John Wiley and Sons, New York, New York (1967).

10. Sharp, G. D., "Electron Microscopy and Viral Particle Function," Transmission of Viruses by the Water Route, Interscience Publishers, a division of John Wiley and Sons, New York, New York, pp. 193-217 (1967).